

سید الرحیم
رحمن



کاربرد داده کاوی در علوم پزشکی

استاد ناظر: جناب آقای دکتر روح الله کلهر

اعضای پنل علمی: جناب آقای دکتر احدی نژاد
جناب آقای دکتر خسروی زاده

ارائه دهنده: علی اسمعیلی

دانشجوی کارشناسی رشته مدیریت خدمات بهداشتی و درمانی



دانشگاه علوم پزشکی قزوین

Conclusion

Result

Methods

Introduction

(۱) در سالیان اخیر شاهد حرکتی مستمر، از تحقیقات صرفاً تئوری به تحقیقات کاربردی به خصوص در زمینه پردازش اطلاعات، برای مسائلی که برای آنها راه حلی موجود نیست و یا براحتی قابل حل نیستند بوده ایم.

(۲) توسعه تئوریک سیستم های دینامیک هوشمند مدل-آزاد که مبتنی بر داده های تجربی برای حل مسائلی که راه حل ندارند.

(۳) شبکه های عصبی مصنوعی جزء این دسته از سیستم های دینامیکی قرار دارند، که با پردازش روی داده های تجربی، دانش یا قانون نهفته در ورای داده ها را به سیستم منتقل می کنند.

(۴) داده کاوی یکی از شاخه های کاربردی هوش مصنوعی است که با جستجوی روابط بین ویژگیها در پایگاههای داده بزرگ الگوهای نهفته را کشف می کند. الگوی کشف شده باید بامعنی بوده و منجر به مزایایی از جمله مزایای اقتصادی شود.

(۵) داده کاوی می تواند الگوهای جدیدی که قبلاً ناشناخته بوده را برای کمک به فرایند تشخیص و درمان استخراج کند.

Conclusion

Result

Methods

Introduction

۶) هدف از داده کاوی در پزشکی ایجاد مدل هایی که به پزشکان در بهبود شیوه های تشخیص، درمان و پیش آگهی کمک کند در سال های اخیر موضوع استفاده از روش های پیشگویانه مانند دسته بندی در تشخیص پزشکی در مقالات پژوهشی رو به افزایش است و پژوهشگران پتانسیل های داده کاوی پیشگویانه را برای استنتاج مدل های مربوط بالینی از داده های بیماران برای فراهم کردن پشتیبانی تصمیم گیری روشن کرده اند.

۷) داده کاوی و کشف دانش می تواند برای خودکارسازی کار تشخیص در پزشکی مورد استفاده قرار گیرد.

۸) از جمله موارد استفاده داده کاوی حل مسائل مربوط به تشخیص بالینی، آنالیز تصاویر پزشکی، پیش بینی بقا و در دامنه وسیعی از زمینه های پزشکی شامل انکولوژی، کاردیولوژی و هماتولوژی، مراقبت های ویژه، تشخیص از روی تصاویر پزشکی، ناباروری، جراحی و ... اشاره کرد.



پیش بینی دیابت ملیتوس نوع دو با استفاده از فنون داده کاوی

* ارائه شده به صورت سخنرانی در دومین همایش فناوری اطلاعات و ارتقا سلامت با محوریت سرطان
دانشگاه علوم پزشکی شهید بهشتی - تهران - شهریور ۹۸



دومین همایش
فناوری اطلاعات و ارتقا سلامت
با محوریت سرطان
شهریور ماه ۱۳۹۸
تهران

دومین همایش
فناوری اطلاعات و ارتقا سلامت
با محوریت سرطان

Second Conference on Information Technology and Health Promotion
Centered on Cancer

روح اله کلههر، سعیده موسوی، نازنین نجفی زادگان ، علی اسمعیلی ، الهام شاه بهرامی ،
اصغر مرتضی قلی*

Conclusion

Result

Methods

Introduction



(۱) دیابت ملیتوس بیماری مزمن است که ۹۰ تا ۹۵ درصد از بیماران دیابتی را در بر می گیرد که بروز و شیوع آن بطور چشمگیری در دهه‌های اخیر بدلیل تغییر در سبک زندگی، رواج چاقی و افزایش طول عمر بیشتر شده است.

(۲) یکی از مشکلات مهم سیستم مراقبت سلامت تشخیص به موقع دیابت نوع دو است و تشخیص زودهنگام و پیشگیری از دیابت، باعث کاهش هزینه و باعث جلوگیری از ایجاد عوارض و بستری شدن بدلیل عوارض شدید این بیماری می شود.

هدف از مطالعه

تشخیص دیر هنگام یا عدم تشخیص دیابت در افراد منجر به گسترش عوارض مختلف عروقی مزمن می شود. از این رو، مطالعه حاضر با هدف بررسی امکان پیش بینی دیابت با استفاده از فنون داده کاوی و ویژگیهای مربوط به ریسک فاکتورهای انجام شد.



Conclusion

Result

Methods

Introduction

این پژوهش از نوع کاربردی و به روش توصیفی- تحلیلی گذشته نگر انجام شد. داده‌های غربالگری دیابت مراجعین به مراکز بهداشتی شهرستان محمدیه به دست آمد. روش‌های مختلفی برای پیاده سازی و اجرای پروژه های داده کاوی وجود دارد. یکی از روش‌های بسیار قوی متدولوژی CRISP (Cross Industry Process For Data Minin) است. در این مقاله مدل پیشنهادی بر اساس CRISP ارائه شده است که شامل پنج فاز است. هر یک از این فازها خود شامل زیر بخشهایی می شوند. حرکت رو به جلو و عقب بین فازهای مختلف نیاز است زیرا ورودی هر فاز به خروجی فاز مرحله قبل وابسته است. هریک از این ۵ فاز در شکل ۱ نشان داده شده است.

Methods

شکل شماره ۱: گام های روش CRISP و مدل پیشنهادی برای انجام پژوهش

گامهای CRISP



مدل پیشنهادی



Conclusion

Result

Methods

Introduction

در این مطالعه تعداد ۱۰۵۵ نمونه وارد مطالعه شدند. از این تعداد از نظر تشخیص دیابت و یا سالم بودن، ۱۵۹ نفر سالم و ۸۹۶ نفر دیابتی تشخیص داده شده بودند.

ابزار گردآوری داده ها، فرم استخراج داده بود که براساس ویژگی های استفاده شده برای غربالگری طراحی شده بود. بر این اساس ویژگی های سن، جنس، فشار خون سیستولیک، فشار خون دیاستولیک، سابقه خانوادگی دیابت، شاخص توده بدنی یا BMI قد، وزن، دور کمر، دور باسن و تشخیص بعنوان ورودی مورد استفاده قرار گرفتند. برای حل مشکل عدم توازن کلاسها از روشهای شناخته شده بازنمونه گیری و روش مبتنی بر الگوریتم استفاده شد.



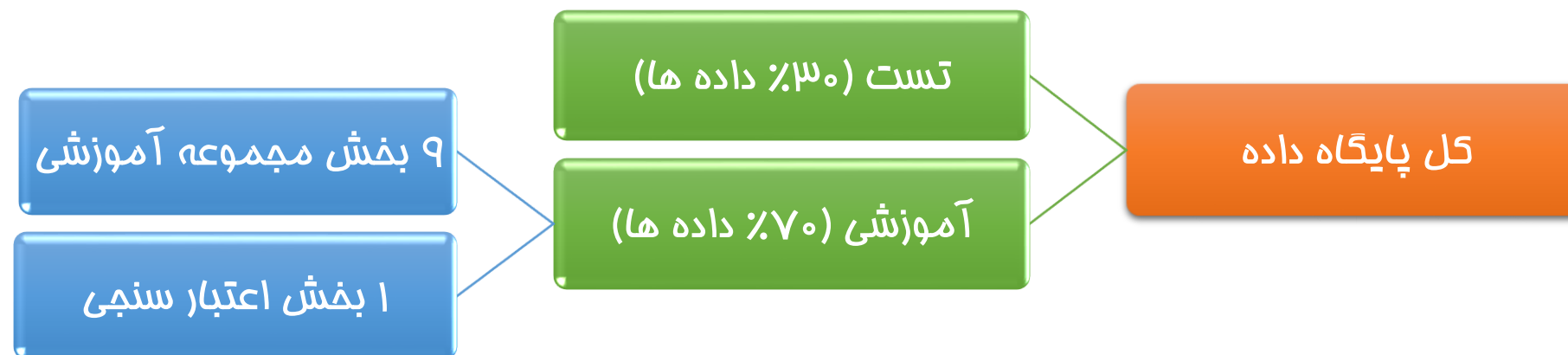
Conclusion

Result

Methods

Introduction

نتایج بدست آمده در این مطالعه بر اساس روش اعتبارسنجی 10-fold cross validation بدست آمده است. (توضیح روش در نمودار زیر)



داده ها در نهایت با استفاده از ۳ روش نزدیک ترین همسایگی، درخت تصمیم گیری و ماشینهای بردار پشتیبان تحلیل و مورد مقایسه قرار گرفتند. جهت تحلیل داده ها از نرم افزار MATLAB ورژن ۲۰۱۳ استفاده شد.

Conclusion

Result

Methods

Introduction

در جدول شماره یک نتایج تمامی ۳ روش کلاس بندی بر اساس چهار معیار ارزیابی ذکر شده مورد مقایسه قرار گرفته است.

بر اساس این جدول در تمامی معیارها، بهترین نتایج توسط درخت تصمیم گیری با صحت (۰.۹۶) و دقت (۰.۸۹) به دست آمده است. پس از آن روشهای نزدیک ترین همسایگی با صحت (۰.۹۶) و دقت (۰.۸۳) و ماشینهای بردار پشتیبان با صحت (۰.۹۴) و دقت (۰.۸۵) قرار داشتند.

جدول شماره ۱: بررسی نتایج روش پیشنهادی به منظور کلاس بندی بیماران دیابتی بر اساس چهار معیار ارزیابی

معیارهای ارزیابی				فنون داده کاوی
صحت (درصد)	نمره F (درصد)	فراخوانی (درصد)	دقت (درصد)	
۰,۹۶	۰,۸۴	۰,۸۶	۰,۸۳	نزدیکترین همسایگی
۰,۹۶	۰,۸۷	۰,۸۶	۰,۸۹	درخت تصمیم گیری
۰,۹۴	۰,۸۲	۰,۸۶	۰,۸۵	ماشینهای بردار پشتیبان

Conclusion

Result

Methods

Introduction

در دومین مرحله از ارزیابی، نتایج تمامی روش های کلاس بندی بر اساس معیار ماتریس اغتشاش مورد مقایسه قرار گرفته اند که نتایج آن در جداول ۲ تا ۴ ارائه شده است. لازم به ذکر است در این جداول تعداد کل نمونه های تست ۳۱۶ نمونه است که از بین تعداد ۲۶۱ نمونه متعلق به کلاس دیابت و ۵۵ نمونه متعلق به کلاس سالم است.

بر اساس این جداول درخت تصمیم گیری بیشترین میزان دقت کلاس بندی را برای هر دو کلاس دیابت و سالم بدست آورده است. یافته های جدول شماره ۲ نشان می دهد که در الگوریتم نزدیکترین همسایه تعداد افراد سالم درست تشخیص داده شده ۴۱ مورد و تعداد افراد دیابتی درست تشخیص داده شده ۲۴۴ مورد می باشد.

جدول شماره ۲: ماتریس اغتشاش مدل ایجاد شده با الگوریتم نزدیکترین همسایه

دیابتی	سالم	
۱۴	۴۱	سالم
۲۴۴	۱۱	دیابتی

Conclusion

Result

Methods

Introduction

یافته های جدول شماره ۳ نشان می دهد که در الگوریتم درخت تصمیم گیری تعداد افراد سالم درست تشخیص داده شده ۴۷ مورد و تعداد افراد دیابتی درست تشخیص داده شده ۲۵۲ مورد می باشد.

جدول شماره ۳: ماتریس اغتشاش مدل ایجاد شده با الگوریتم درخت تصمیم گیری

	سالم	دیابتی
سالم	۴۷	۸
دیابتی	۹	۲۵۲

یافته های جدول شماره ۴ نشان می دهد که در الگوریتم ماشین بردار پشتیبان تعداد افراد سالم درست تشخیص داده شده ۴۵ مورد و تعداد افراد دیابتی درست تشخیص داده شده ۲۵۰ مورد می باشد.

جدول شماره ۴: ماتریس اغتشاش مدل ایجاد شده با الگوریتم ماشین بردار پشتیبان

	سالم	دیابتی
سالم	۴۵	۱۰
دیابتی	۱۱	۲۵۰

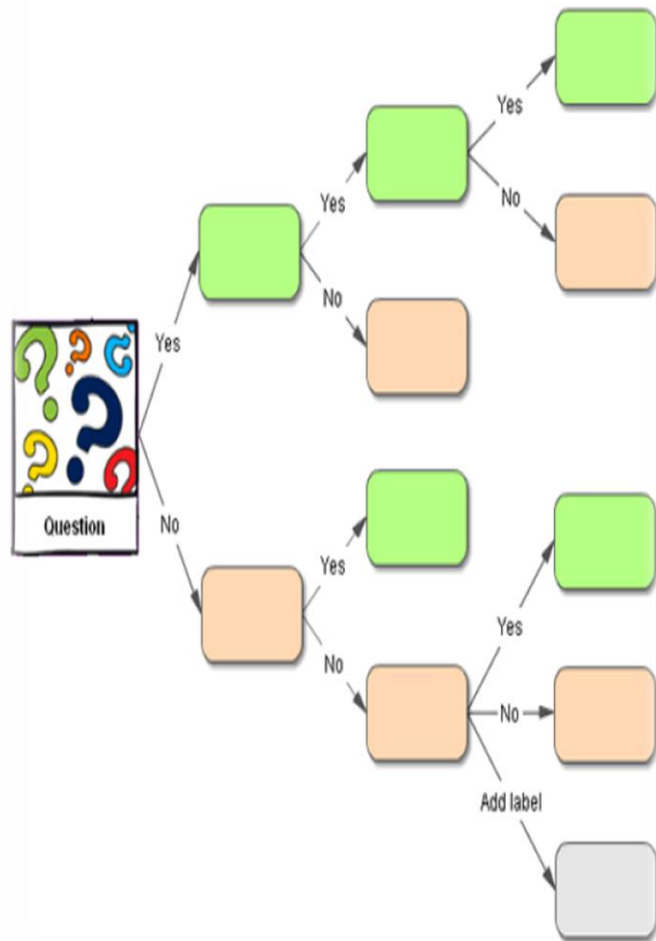
Conclusion

Result

Methods

Introduction

Decision Tree



یافته های دیگر مطالعات نتایج متفاوتی نشان داده است. حبیبی در مطالعه خود با عنوان "مطالعه مدل‌های پیش بینی دیابت ملیتوس نوع دو با استفاده از فنون داده کاوی برای طراحی سیستم خبره" به این نتیجه رسید که در مقایسه مدل‌های پیش بینی بالاترین اندازه F مربوط به الگوریتم $J48$ درخت تصمیم بود که با استفاده از بوستینگ بدست آمد، بعد از آن هزینه- حساسیت با اندازه F برابر با $۴۹۲ / ۰$ از بقیه روشها عملکرد بهتری داشت.

مطالعه هوانگ و همکاران که هدفشان ساخت مدل دسته بندی بر روی داده های دیابت نوع ۲ بود تایید کننده عملکرد بهتر مدل درخت تصمیم در این دسته بندی بود. نتایج این دو مطالعه از نظر تعیین بالاتری دقت روش داده کاوی یافته های مطالعه ما را تایید می کرد.

Conclusion

Result

Methods

Introduction

با اینحال پژوهشهایی نیز بر توانایی بالاتر سایر فنون نسبت به درخت تصمیم تاکید کردند از جمله آنها آرونا در مطالعه مقایسه ای فنون داده کاوی که بر روی داده های Pima Indians با استفاده از نرم افزار WEKA انجام دادند بالاترین دقت و حساسیت را مربوط به ماشین بردار پشتیبان ذکر کردند (به ترتیب ۵/۹۶ و ۶/۹۸ درصد) و بعد از آن بیز ساده و سپس درخت تصمیم با الگوریتم J48 قرار گرفت.

همچنین جیتراکول و همکاران در مقایسه کارایی دو رویکرد نمونه گیری کاهشی و افزایشی اسموت بر روی داده های نامتوازن با استفاد از فنون شبکه عصبی، ماشین بردار پشتیبان و نزدیکترین همسایگی بر روی داده های نامتوازن از چهار مجموعه داده موجود در مخزن داده یو سی آی استفاده کردند. آنها دریافتند با استفاده از اسموت مدل شبکه عصبی نسبت به بقیه مدلها صحت بهتری داشت.

به نظر می رسد این اختلاف در نتایج مطالعات مختلف شاید ناشی از تفاوت در داده های مورد استفاده هر مطالعه باشد و همین داده های متنوع موجب عملکرد متفاوت مدلهای مختلف شده است. بنابراین هم داده های مورد استفاده و هم ویژگیهای استفاده شده در مدل در نتیجه نهایی مطالعه موثر است.

Conclusion

Result

Methods

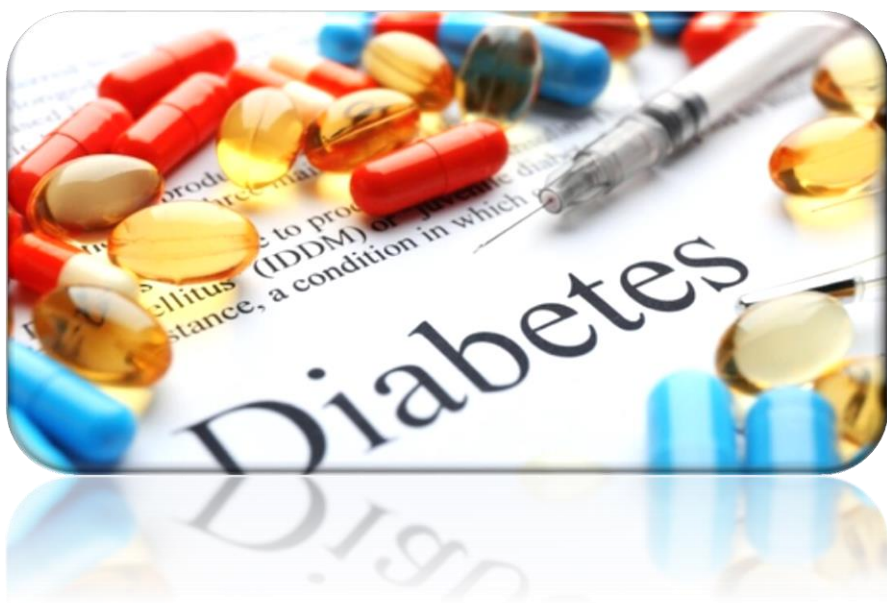
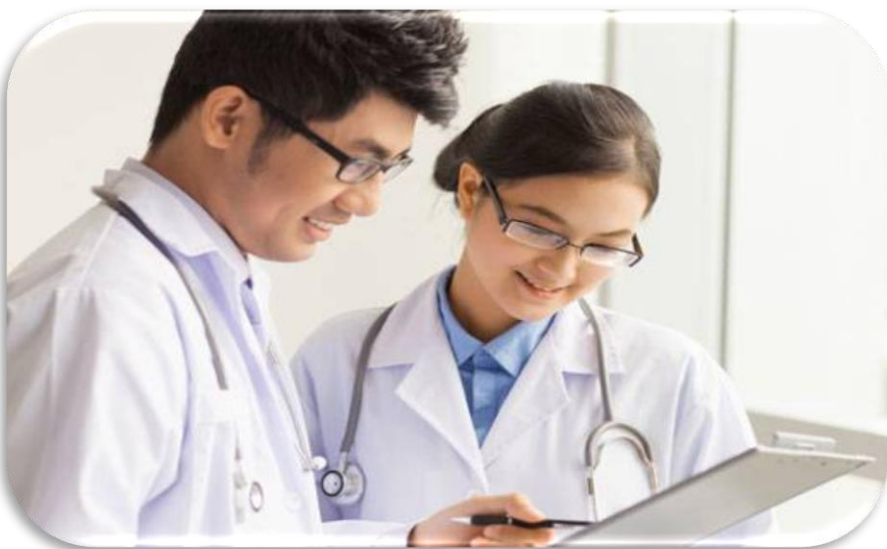
Introduction

نتیجه گیری

بر اساس نتایج ارائه شده، مدل درخت تصمیم گیری با صحت (۰.۹۶) و دقت (۰.۸۹) بهترین نتایج را در کلاس بندی نمونه های تست نشان می دهد. از موارد کاربرد نتایج این پژوهش استفاده از این مدلها در آزمایشگاه ها برای افرادی که آزمایشهای متنوعی داده اند به منظور پیش بینی میزان قندخون و تشخیص دیابت می باشد.

پیشنهادهات

- از این مدلها می توان به عنوان دستیار پزشک در بیمارستان های آموزشی درمانی که عملیات تشخیص توسط دانشجویان رشته پزشکی انجام می شود استفاده کرد.
- بکار گیری روش های داده کاوی در جهت استنتاج مدل هایی برای پیش بینی و کمک به تشخیص بیماری دیابت نوع ۲



1. Tuomilehto J, Lindström J, Eriksson JG, Valle TT, Hämäläinen H, Ilanne-Parikka P, et al. Prevention of Type 2 Diabetes Mellitus by Changes in Lifestyle among Subjects with Impaired Glucose Tolerance. *New England Journal of Medicine*2001;344(18):1343-50.
2. American Diabetes Association. Standards of medical care in diabetes 2012. *Diabetes Care*2012 Jan;35 Suppl 1:11-63.
3. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care*2013;36(Supplement 1):S67-S74.
4. Heydari I, Radi V, Razmjou S, Amiri A. Chronic complications of diabetes mellitus in newly diagnosed patients. *International Journal of Diabetes Mellitus*2010;2(1):61-3.
5. Karter AJ, Stevens MR, Herman WH, Ettner S, Marrero DG, Safford MM, et al. Out-of-Pocket Costs and Diabetes Preventive Services The Translating Research Into Action for Diabetes (TRIAD) study. *Diabetes Care*2003;26(8):2294-9.
6. Luijckx H, Schermer T, Bor H, van Weel C, Lagro-Janssen T, Biermans M, et al. Prevalence and incidence density rates of chronic comorbidity in type 2 diabetes patients: an exploratory cohort study. *BMC medicine*2012;10(1):128.
7. American Diabetes Association. Standards of medical care in diabetes--2012. *Diabetes Care*2012 Jan;35 Suppl 1:S11-63.
8. American Diabetes Association. Screening for type 2 diabetes. *Diabetes Care*2004;27:S11.
9. Beagley J, Guariguata L, Weil C, Motala AA. Global estimates of undiagnosed diabetes in adults. *Diabetes research and clinical practice*2014;103(2):150-60.
10. Gregg EW, Geiss LS, Saaddine J, Fagot-Campagna A, Beckles G, Parker C, et al. Use of diabetes preventive care and complications risk in two African-American communities. *American journal of preventive medicine*2001;21(3):197-202.
11. Zhuo X, Zhang P, Hoerger TJ. Lifetime direct medical costs of treating type 2 diabetes and diabetic complications. *American journal of preventive medicine*2013;45(3):253-61.
12. Rafeh R, Arbabi M. Data Mining Techniques to Diagnose Diabetes Using Blood Lipids. *Scientific Journal of Ilam University of Medical Sciences* 2015; 23(4): 239-47. [in Persian]
13. Witten IH, Frank E. *Data Mining: Practical machine learning tools and techniques*: Morgan Kaufmann; 2005.
14. Zhao J, Wang T, editors. A general framework for medical data mining. *Future Information Technology and Management Engineering (FITME)*, 2010 International Conference on; 2010: IEEE.
15. Patil BM, Joshi RC, Toshniwal D. Hybrid prediction model for Type-2 diabetic patients. *Expert Systems with Applications*2010;37(12):8102-8.
16. Soni J, Ansari U, Sharma D, Soni S. Predictive data mining for medical diagnosis: An overview of heart disease prediction. *International Journal of Computer Applications*2011;17(8):43-8.
17. Rajalakshmi K, Chandra Mohan S, Dhinesh BS. Decision Support System in Healthcare Industry. *International Journal of Computer Applications*2011;26(9):42-4.
18. Patil B, Joshi R, Toshniwal D, editors. Association rule for classification of type-2 diabetic patients. *Machine Learning and Computing (ICMLC)*, 2010 Second International Conference on; 2010: IEEE.
19. Temurtas H, Yumusak N, Temurtas F. A comparative study on diabetes disease diagnosis using neural networks. *Expert Systems with Applications*2009;36(4):8610-5.
20. Habibi S, Ahmadi M, Alizadeh S. Type 2 Diabetes Mellitus Screening and Risk Factors Using Decision Tree: Results of Data Mining. *Global Journal of Health Science*. 2015;7(5):304-310. doi:10.5539/gjhs.v7n5p304.
21. Alizadeh S, Ghazanfari M, Teimorpour B .Data Mining and Knowledge Discovery, Publication of Iran University of Science and Technology .2nd ed. 2011. [in Persian]
22. Habibi S. A study on diabetes type II predictive models applying data mining techniques in expert systems development [dissertation]. Tehran: Iran University of Medical Sciences; School of Health Management and Information Science: 2015. [in Persian]
23. Huang Y, McCullagh P, Black N, Harper R. Feature selection and classification model construction on type 2 diabetic patients' data. *Artificial Intelligence in Medicine*2007;41(3):251-62
24. Aruna S, Rajagopalan S, Nandakishore L. An Empirical Comparasion of Supervised Learning Algorithms in Disease Detection. *International Journal of Information Technology Convergence and Services–IJITCS*2011;1:81-92.
25. Jeatrakul P, Wong KW, Fung CC. Classification of imbalanced data by combining the complementary neural network and SMOTE algorithm. *Neural Information Processing Models and Applications*: Springer; 2010. 152-9.
26. Tama BA, Rodiyatul F, Hermansyah H. An Early Detection Method of Type-2 Diabetes Mellitus in Public Hospital. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*2013;9(2):287-94.

تشخیص سرطان معده با استفاده از روش های داده کاوی

* ارائه شده به صورت پوستر در دومین همایش فناوری اطلاعات و ارتقا سلامت با محوریت سرطان

دانشگاه علوم پزشکی شهید بهشتی - تهران - شهریور ۹۸



دومین همایش
فناوری اطلاعات و ارتقا سلامت
با محوریت سرطان
شهریور ماه ۱۳۹۸
تهران

دومین همایش
فناوری اطلاعات و ارتقا سلامت
با محوریت سرطان

Second Conference on Information Technology and Health Promotion
Centered on Cancer

اصغر مرتضی قلی، روح اله کلهر*، سعیده موسوی، عبدالله کشاورز، علی اسمعیلی، الهام
شاه بهرامی

Conclusion

Result

Methods

Introduction

(۱) سرطان معده یکی از شایع ترین نوع سرطان ها می باشد که تشخیص آن در مراحل اولیه بدلیل نداشتن نشانه و علائم خاص سخت می باشد با این وجود دانشمندان، بعضی از عوامل خطر سازی که فرد را مستعد پیشرفت به سمت سرطان معده می کنند را شناسایی کرده اند. بعضی از این عوامل عبارتند از: عفونت باکتریای، جنسیت، سن، نژاد، محیط و محل زندگی، رژیم غذایی، عمل جراحی قبلی روی معده، و کم خونی بدخیم. همچنین گروه خونی نیز از عوامل موثر شناخته شده است. گرچه شیوع جهانی سرطان معده رو به کاهش است ولی درصد شیوع آن در کشورهای آسیایی همچنان بالا است

(۲) این سرطان بیشتر در افراد بالای ۴۰ سال رایج است که شیوع آن در بین مردان دو برابر زنان می باشد. سرطان معده پس از سرطان های ریه، پستان و روده چهارمین سرطان شایع در جهان می باشد. بطور کلی سرطان معده رتبه دوم در مرگ و میر ناشی از سرطان را در جهان دارا می باشد. متأسفانه تا وقتی که این سرطان به سطح پیشرفته خود نرسیده باشد هیچ علائمی از خود نشان نمی دهد، عمدتاً برای تشخیص سرطان معده از روشهایی همچون: بیوپسی، آندوسکپی، سونوگرافی، سی تی اسکن، اشعه ایکس رادیو گرافی و تست های بالینی مخصوص استفاده می شود

هدف از مطالعه

هدف از این مطالعه امکان پیش بینی و تشخیص سرطان معده با استفاده از روش های داده کاوی و ویژگیهای مربوط به ریسک فاکتورهای بیماری می باشد.



Conclusion

Result

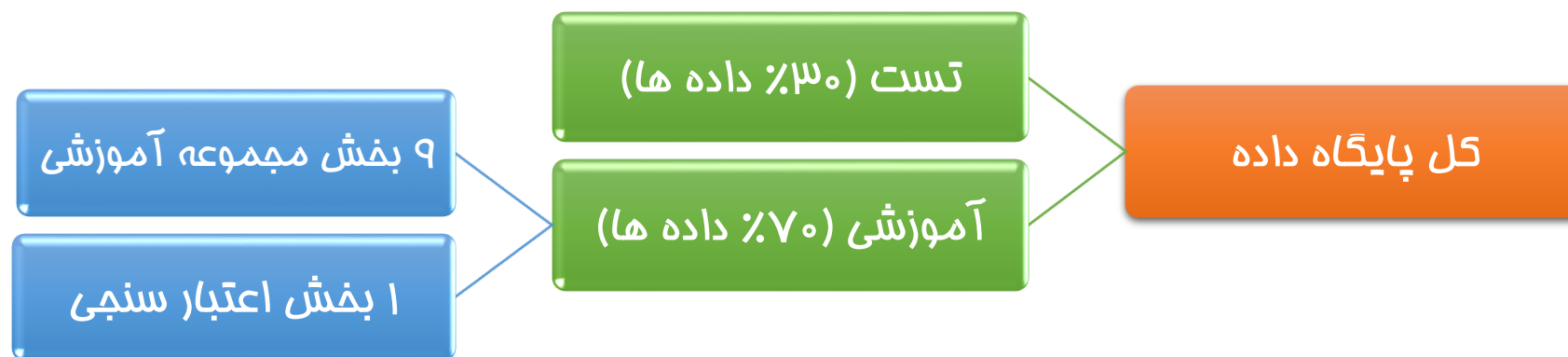
Methods

Introduction

این پژوهش از نوع کاربردی و به روش توصیفی- تحلیلی گذشته نگر است که در سال ۲۰۱۷ دو بیمارستان رجایی و بوعلی انتخاب شدند با تعداد ۴۰۵ نمونه از دو کلاس بیمار و سالم. انجام شده است.

تعداد ویژگی های موجود در پایگاه داده ۱۱ ویژگی بود که عبارت بودند از جنسیت، سن، کاهش وزن، درد شکم، استفراغ، بی اشتهایی، مشکل در بلعیدن غذا، کم خونی حاد، مدفوع تیره و یا آغشته به خون، و توده شکمی.

ابزار گردآوری داده ها، فرم استخراج داده بود که براساس ویژگی های استفاده شده برای غربالگری طراحی شده بود. برای حل مشکل عدم توازن کلاسها از روشهای شناخته شده بازنمونه گیری و روش مبتنی بر الگوریتم استفاده شد. نتایج بدست آمده در این مطالعه بر اساس روش اعتبارسنجی 10-fold cross validation بدست آمده است.



Conclusion

Result

Methods

Introduction

در این مطالعه از روش های یادگیری ماشین استفاده شده است. دلیل استفاده از این روش ها این است که در یک کاربرد پزشکی، داده های (اطلاعات) پزشکی بیمار معمولاً با تعداد بسیار زیاد فاکتور (ویژگی) همراه است که در این صورت لحاظ نمودن تمامی این ویژگی ها در زمان تصمیم گیری در مورد وضعیت بیمار توسط پزشک و یا با استفاده از روش های ریاضیاتی ممکن است با خطا و پیچیدگی همراه بوده و در نتیجه عملکرد بالایی را ارائه نخواهد کرد. این مسئله در بیماران سرطان معده با وجود تعداد بسیار ویژگی از اهمیت ویژه ای برخوردار است. برای همین منظور در این پژوهش از چهار روش یادگیری ماشین بردار پشتیبان، درخت تصمیم، مدل بیزین و الگوریتم نزدیک ترین همسایگی به منظور طبقه بندی بیماران سرطان معده استفاده شده است تا با بررسی این چهار روش تعیین شود که کدام یک از روش های یادگیری ماشین دقت بالاتری را نسبت به سایر روش ها در طبقه بندی نمونه های سرطان معده بدست می آورد.

به منظور ارزیابی مدل بر روی پایگاه داده بیماران سرطان معده، معیارهای ارزیابی عبارتند از Recall، Precision، F-score، Accuracy و Confusion Matrix (CM) جهت تحلیل داده ها از نرم افزار MATLAB ورژن ۲۰۱۳ استفاده شد.

Conclusion

Result

Methods

Introduction

نتایج تمامی چهار روش کلاس‌بندی بر اساس چهار معیار ارزیابی ذکر شده مورد مقایسه قرار گرفته است (جدول زیر). برای این اساس در تمامی معیارها، بهترین نتایج توسط ماشین بردار پشتیبان با صحت (90.08) و دقت (90.78) بدست آمده است. پس از روش پیشنهادی به ترتیب درخت تصمیم، مدل بیزین و الگوریتم نزدیک‌ترین همسایگی در رتبه‌های بعدی کلاس‌بندی بیماران سرطان معده قرار گرفته‌اند.

Method	performance measure			
	F-score	Precision	Recall	Accuracy
k-NN	87.17	89.47	85	87.60
NB	87.99	84.61	91.66	87.60
DT	88.37	87.69	89.06	87.89
SVM	91.99	90.78	93.24	90.08

Result

در دومین مرحله از ارزیابی، نتایج تمامی روش های کلاس بندی بر اساس معیار ماتریس CM مورد مقایسه قرار گرفته اند که نتایج آن در جدول رو به رو ارائه شده است. لازم به ذکر است در جدول رو به رو تعداد کل نمونه های تست ۱۲۲ نمونه است که از بین تعداد ۶۲ نمونه متعلق به کلاس سرطان و ۶۰ نمونه متعلق به کلاس سالم است. بر اساس این جداول ماشین بردار پشتیبان بیشترین میزان دقت کلاس بندی را برای هر دو کلاس سرطانی و سالم بدست آورده است.

نتایج کلاس بندی بر اساس معیار ماتریس CM			روش
Cancer	Healthy	<div>Predicted</div> <div>Actual</div>	ماشین بردار پشتیبان
1	57	Healthy	
61	3	Cancer	
Cancer	Healthy	<div>Predicted</div> <div>Actual</div>	درخت تصمیم
5	51	Healthy	
57	9	Cancer	
Cancer	Healthy	<div>Predicted</div> <div>Actual</div>	بیزین
3	53	Healthy	
59	7	Cancer	
Cancer	Healthy	<div>Predicted</div> <div>Actual</div>	الگوریتم نزدیکترین همسایگی
8	49	Healthy	
54	11	Cancer	

Conclusion

Result

Methods

Introduction

بر اساس نتایج ارائه شده ماشین بردار پشتیبان به دلیل انعطاف پذیری بالا در یادگیری مسائل مختلف و همچنین به دلیل بهره-گیری از قدرت تعمیم پذیری بالا، عمل کلاس بندی نمونه ها را با دقت بالاتری نسبت به سایر روش های کلاس بندی انجام داده و به همین دلیل بهترین نتایج را در کلاس بندی نمونه ها بر روی تمامی معیارها بدست آورده است. لازم به ذکر است که درخت تصمیم به دلیل وابستگی به داده های آموزشی و قدرت تعمیم پذیری کمتر نسب به روش ماشین بردار پشتیبان، در رتبه دوم روش های کلاس بندی قرار گرفته است. روش بیزین نیز به دلیل استفاده از فرض استقلال ویژگی ها از یکدیگر، کیفیت هر ویژگی را در کلاس بندی نمونه ها مستقل از سایر ویژگی در نظر می گیرد این در حالی است که در بسیاری از حالات یکی ویژگی ممکن است به تنهایی تاثیری در کلاس بندی نمونه ها نداشته باشد اما زمانی که با سایر ویژگی ها ترکیب می شود، نتایج بهتری را ارائه دهد. الگوریتم نزدیک ترین همسایگی نیز به دلیل عدم استفاده تکنیک یادگیری و کلاس بندی نمونه ها تنها براساس معیار فاصله و همچنین به دلیل حساسیت به نویز، بدترین نتایج کلاس بندی نمونه های سرطانی را نسبت به سایر روش ها بدست آورده است.

Conclusion

Result

Methods

Introduction

بر اساس نتایج ارائه شده ، ماشین بردار پشتیبان بهترین نتایج را در کلاس بندی نمونه ها بر اساس معیارهای ارزیابی ذکر شده و مقایسه با سایر روش های یادگیری ماشین نشان می دهد. همسو با نتایج پژوهش مطالعه Wang و همکاران در مراکز درمانی چین نشان می دهد که الگوریتم ماشین بردار پشتیبان روش مناسبی بر اساس معیار های ارزیابی برای تشخیص سرطان معده دارد و دقت طبقه بندی SVM به ویژه در ۴ نوع طبقه بندی بسیار بالا است مقدار دقت آن ۹۱/۷ است که نشان دهنده کاربرد بالقوه مدل تشخیصی SVM در تشخیص سرطان است.

این در حالی است که در مطالعه محمودی در یکی از مراکز درمانی تبریز، الگوریتم درخت تصمیم روش مناسبی برای شناسایی عوامل موثر بر بروز سرطان معده می باشد. دقت پیش بینی به دست آمده از الگوریتم درخت تصمیم برابر ۵۶/۸۵ به دست آمد که نتیجه ی بسیار خوبی در پیش بینی سرطان معده است (۱۳). همچنین در مطالعه ی silvera نتایج بیانگر کاربرد روش درخت تصمیم در پیش بینی عوامل موثر بروز سرطان معده است و رفالکس معده، مهم ترین عامل بروز سرطان معده است.

Conclusion

Result

Methods

Introduction

از سوی دیگر در مطالعه kirshners و همکاران یک روش چند لایه با ترکیب تکنیک های خوشه بندی و درخت تصمیم یک سیستم تشخیص و پیش بینی خطر ابتلا به سرطان سرطان معده ارائه کرده اند که نتایج حاصل از درخت تصمیم در این مطالعه نشان می دهد که جنسیت افراد، خطرات شغلی، کاهش وزن، سابقه ی خانوادگی، مصرف الکل، درد شکم همراه با خون در مدفوع از عوامل و عالیم سرطان معده است. معیار دقت در خصوص کاربرد الگوریتم SVM در مطالعه حاضر نزدیک ۹۱ درصد بوده است.

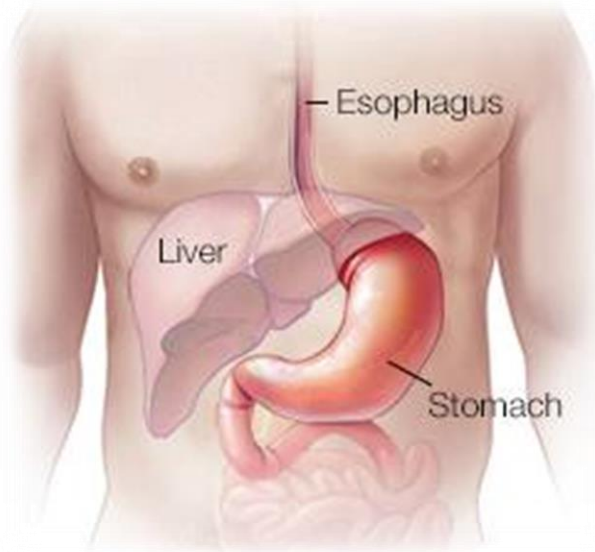
در این خصوص مطالعه ی احمد زاده و همکاران نشان داد که سیستم پیشنهادی با استفاده از ترکیب روش های مختلف موفق شد با تکیه بر ویژگی های پایگاه داده در قالب ترکیب و تعامل به دقت شناسایی ۸۵.۸٪ دست یابد این درحالی است که روش های حاضر علی رغم دقت بالا، هزینه بر و وقت گیر می باشند.

Conclusion

Result

Methods

Introduction



ماشین بردار پشتیبان بهترین نتایج را در کلاس بندی نمونه ها بر اساس معیارهای ارزیابی ذکر شده و مقابسه با سایر روش های یادگیری ماشین نشان می دهد. در ادامه ی این مطالعه می توان روش های انتخاب ویژگی را به منظور بهبود نتایج کلاس بندی به روش پیشنهادی افزود. علاوه بر این می توان عملیات انتخاب ویژگی را با استفاده از روش های تکاملی و تاثیر آن در مسئله کلاس-بندی بیماران سرطانی مطالعه کرد. همچنین با توجه ویژگی عدم تعادل نمونه های یک پایگاه داده پزشکی، مسئله متعادل سازی مجموعه داده بیماران سرطان معده و تاثیر آن در بهبود نتایج کلاس بندی چهار روش کلاس بندی مورد استفاده در این پژوهش مورد مطالعه قرار داد. علاوه بر این می توان از روش های تکاملی به منظور یافتن مقادیر بهینه الگوریتم ماشین بردار پشتیبان بهره برد تا پیچیدگی محاسباتی آن را کاهش داد.

1. Kim H-I, Park M, Song K, Woo Y, Hyung W. Rapid and safe learning of robotic gastrectomy for gastric cancer: multidimensional analysis in a comparison with laparoscopic gastrectomy. *European Journal of Surgical Oncology*. 2014;40(10):1346-54.
2. de Castro Mayor J, Caño JV, Aragón JC, Andrés GB, Herranz FA, Hernández CF. Locally advanced prostate cancer. Definition, diagnosis and treatment. *Archivos españoles de urologia*. 2018;71(3):231-8.
3. Bartfeld S, Bayram T, van de Wetering M, Huch M, Begthel H, Kujala P, et al. In vitro expansion of human gastric epithelial stem cells and their responses to bacterial infection. *Gastroenterology*. 2015;148(1):126-36. e6.
4. Shimada S, Sawada N, Ishiyama Y, Nakahara K, Maeda C, Mukai S, et al. Impact of obesity on short-and long-term outcomes of laparoscopy assisted distal gastrectomy for gastric cancer. *Surgical endoscopy*. 2018;32(1):358-66.
5. Alkebsi L, Ideno Y, Lee J-S, Suzuki S, Nakajima-Shimada J, Ohnishi H, et al. Gastroduodenal Ulcers and ABO Blood Group: the Japan Nurses' Health Study (JNHS). *Journal of epidemiology*. 2018;28(1):34-40.
6. Zhu L, Luo W, Su M, Wei H, Wei J, Zhang X, et al. Comparison between artificial neural network and Cox regression model in predicting the survival rate of gastric cancer patients. *Biomedical reports*. 2013;1(5):757-60.
7. Billiar T, Andersen D, Hunter J, Brunicardi F, Dunn D, Pollock RE, et al. *Schwartz's Principles of Surgery*. McGraw-Hill Professional; 2009.
8. Afshar S, Abdolrahmani F, Tanha FV, Seaf MZ, Taheri K, editors. Quick and reliable diagnosis of stomach cancer by artificial neural network. *Proceedings of the 10th WSEAS international conference on Mathematics and computers in biology and chemistry*; 2009: World Scientific and Engineering Academy and Society (WSEAS).
9. Biglarian A, Hajizadeh E, Kazemnejad A, Zali M. Application of artificial neural network in predicting the survival rate of gastric cancer patients. *Iranian journal of public health*. 2011;40(2):80.
10. Menhaj M. Fundamentals of neural networks. *Computational intelligence*. 1998;1(1).
11. Kurt I, Ture M, Kurum AT. Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. *Expert systems with applications*. 2008;34(1):366-74.
12. Wang H, Huang G. Application of support vector machine in cancer diagnosis. *Medical Oncology*. 2011;28(1):613-8.
13. Mahmoodi SA, Mirzaie K, Mahmoodi SM. Determining the Effective Factors in the Incidence of Gastric Cancer by Using Data Mining Approach. *Payavard Salamat*. 2017;11(3):332-41.
14. Silvera SAN, Mayne ST, Gammon MD, Vaughan TL, Chow W-H, Dubin JA, et al. Diet and lifestyle factors and risk of subtypes of esophageal and gastric cancers: classification tree analysis. *Annals of epidemiology*. ۲۰۰۰;(۱)۲۴;۲۰۱۴ .
15. Kirshners A, Polaka I, Aleksejeva L. Gastric cancer risk analysis in unhealthy habits data with classification algorithms. *Information Technology and Management Science*. 2015;18(1):97-102.
16. Ahmadzadeh D, Fiyuzi M. Providing an intelligent system for diagnosis of gastric cancer disease based on a backup vector machine classification algorithm 2013 [Available from: https://www.civilica.com/Paper-AISST01-AISST01_183.html].
17. Wang L. *Support vector machines: theory and applications*: Springer Science & Business Media; 2005.
18. Zhang Y, Lai X, Zeng Q, Li L, Lin L, Li S, et al. Classifying low-grade and high-grade bladder cancer using label-free serum surface-enhanced Raman spectroscopy and support vector machine. *Laser Physics*. 2018;28. ۰۳۰۶.۰۳:(۳)
19. Chang R-F, Wu W-J, Moon WK, Chou Y-H, Chen D-R. Support vector machines for diagnosis of breast tumors on US images. *Academic radiology*. 2003;10(2):189-97.
20. Jumutc V, Zayakin P, Borisov A. Ranking-based kernels in applied biomedical diagnostics using a support vector machine. *International Journal of Neural Systems*. 2011;21(06):459-73.
21. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Machine learning*. 2002;46(1-3):389-4. ۲۲
22. Abe S. *Support vector machines for pattern classification*: Springer; 2005.
23. Yoo I, Alafaireet P, Marinov M, Pena-Hernandez K, Gopidi R, Chang J-F, et al. Data mining in healthcare and biomedicine: a survey of the literature. *Journal of medical systems*. 2012;36(4):2431-48.
24. Moulin L, Da Silva AA, El-Sharkawi M, Marks RJ. Support vector machines for transient stability analysis of large-scale power systems. *IEEE Transactions on Power Systems*. 2004;19(2):818-25.

با شکر از تو پیچی که
منذول داسید.

